

Mini- and Microsatellites

Claes Ramel

Department of Genetic and Cellular Toxicology, Stockholm University, Stockholm, Sweden

While the faithful transmission of genetic information requires a fidelity and stability of DNA that is involved in translation into proteins, it has become evident that a large part of noncoding DNA is organized in repeated sequences, which often exhibit a pronounced instability and dynamics. This applies both to longer repeated sequences, minisatellites (about 10–100 base pairs), and microsatellites (mostly 2–4 base pairs). Although these satellite DNAs are abundantly distributed in all kinds of organisms, no clear function has been discerned for them. However, extension of trinucleotide microsatellite sequences has been associated with several severe human disorders, such as Fragile X syndrome and Huntington's disease. Rare alleles of a minisatellite sequence have been reported to be associated with the *ras* oncogene leading to an increased risk for several human cancers. A dynamic behavior of repeated DNA sequences also applies to telomeres, constituting the ends of the chromosomes. Repeated DNA sequences protect the chromosome ends from losing coding sequences at cell divisions. The telomeres are maintained by the enzyme telomerase. Somatic cells, however, lose telomerase function and gradually die. Cancer cells have activated telomerase and therefore they acquire immortality. — *Environ Health Perspect* 105(Suppl 4):781–789 (1997)

Key words: minisatellites, microsatellites, amplification, genetic instability, fingerprinting, transpositions, recombination, mismatch repair, telomere

Introduction

Since around 1970 our concept of DNA has undergone a pronounced modification in two respects: the stability of DNA and the organization of the genetic material in living organisms. Previously DNA was considered a highly stable entity, which was subjected only to alterations through rare mutational events. This notion was supported by experimental evidence in the 1950s, which showed that DNA replication proceeds with an extremely low frequency of errors. The stability and precision of the genetic system were manifested by the

formulation of the “Central Dogma,” which holds that the flow of information in the cells proceeds from DNA to RNA and from RNA to proteins.

In the 1970s it was found that DNA was far more dynamic than anticipated and that the central dogma was not invariably correct. The detection of reverse transcriptase showed that RNA could be transcribed to DNA; this had important consequences for many cellular processes, such as the insertion of mobile elements, the foundation of pseudogenes from mRNA, and retroviral replication.

The other area of DNA research, for which the last two decades have provided a fundamentally new concept of DNA, concerns the organization of the genetic material in higher organisms. The vast majority of DNA—about 97% in human DNA—does not give rise to any proteins. The functioning of all this DNA has been obscure from the beginning and it was named “selfish DNA” by Orgel and Crick (1). Other names, such as junk DNA, parasite DNA, and extra DNA illustrate the confusion about the functioning of this DNA. Some organisms have a remarkably high amount of DNA. Thus amphibians can have up to 20 times more DNA than man (1). The reason for this spectacular variation

in DNA content between organisms is still obscure.

Some of the noncoding DNA occurs as introns, which are spliced away before translation, but most of it is organized as repeated sequences, which somewhat constitute “biological dynamite,” in the sense that they are apt to exhibit a high degree of instability and thus are responsible for much of the instability of DNA mentioned above.

Research in more recent years on repeated sequences of DNA has given an insight into the behavior and biological consequences of changes in these units. Alterations, particularly of microsatellites, have been shown to be connected with several severe human disorders. Although it appears that these repeated sequences do not provide any obvious benefit to the organism, the fact that their alteration can have severe effects nevertheless indicates some kind of a function behind the occurrence of these seemingly nonsense DNA sequences. In the present report an attempt will be made to summarize current knowledge of mini- and microsatellites as well as some other repeated sequences of DNA.

Major Types of Repeated DNA Sequences

Amplification of DNA is not restricted to noncoding sequences; many coding genes are also amplified or can go through such a process. Amplification of nuclear oncogenes, leading to an increase of their expression, is connected with the induction of cancer. In many cases an amplification of genes constitutes a defense mechanism for detoxication of exogenous agents, i.e., amplification of metallothioneins against heavy metals and the amplification up to 1000 times of the dehydrofolate gene giving resistance to methotrexate (2). In *Drosophila*, a specific amplification controlling element is responsible for the amplification of corion genes during the early development stages of the embryo.

The ribosomal DNA in the proximal heterochromatin of *Drosophila* occurs as a highly amplified gene, and the optimal number of genes is gradually restored in case of a deletion of a part of the heterochromatin.

Concerning shorter and mostly noncoding repeats of DNA, which are the main subject of this presentation, we can recognize two classes—minisatellites: up to 100 bp, but mostly about 9 to 30 bp; and microsatellites: 2 to 4 bp, telomeres and telomericlike sequences, and centromeres.

This paper was prepared as background for the Workshop on Susceptibility to Environmental Hazards convened by the Scientific Group on Methodologies for the Safety Evaluation of Chemicals (SGOMSEC) held 17–22 March 1996 in Espoo, Finland. Manuscript received at EHP 5 November 1996; accepted 18 November 1996.

Address correspondence to Dr. C. Ramel, Department of Genetic and Cellular Toxicology, University of Stockholm, S-106 91 Stockholm, Sweden. Telephone: 46 8 16 20 51. Fax: 46 8 612 40 04. E-mail: claes.ramel@genetics.su.se

Abbreviations used: AR, androgen receptor; HNPCC, hereditary nonpolyposis colon cancer; IDDM, type I diabetes mellitus; IGM, immunoglobulin heavy-chain gene; INS, insulin gene; LINE, long interspersed element; MVR, minisatellite variant repeat; PCR, polymerase chain reaction; RAP, repeat associated point mutations; RIP, repeat-induced point mutations; VNTR, variable number of tandem repeats.

Minisatellites

Occurrence. Minisatellites are regions of the genome with noncoding, tandemly repeated sequences of up to about 100 bp (3,4). The number of minisatellite loci in the human genome has been estimated to be 1500 per haploid genome (5,6). Many of these loci exhibit an extreme polymorphism due to variation in the number of repeats, called variable number of tandem repeats (VNTR). The background of this genetic variability is a mutation rate that can exceed 10% per gamete (7). Jeffreys et al. (3) detected and developed DNA probes that are able to simultaneously detect large numbers of hypervariable minisatellite loci. Hybridization to digested and electrophoresed DNA with these core sequences at low stringency detects a pattern of fragments that is unique for unrelated individuals. These properties of minisatellites provide the background for the "fingerprint analyses" (8), which have found several important applications, including as a powerful tool in forensic medicine, as markers for linkage studies in genetic analyses, and as a means for establishing kinship between individuals, including paternity determination. Through the development of a system of polymerase chain reaction (PCR) (minisatellite variant repeat [MVR], below) by Jeffreys and co-workers, it is possible to analyze single pairs of minisatellite alleles (9). This has enabled a measurement of changes in the number of repeats and also the occurrence, frequency and location of point mutations along the sequence of repeats in a single allele. This development has been important to the study of the mechanism for genetic changes of these repeated sequences and the genetic instability involved.

The minisatellites of the human genome are not evenly distributed but are primarily localized at the ends of the chromosomes, which implies a limitation in the use of these sequences in linkage analyses (10). This subtelomeric localization of the human minisatellites is correlated with a high density of chiasmata during meiosis, indicating an association with meiotic crossing over (11,12). The human X chromosome has few minisatellites, but there is a cluster of minisatellites in the X-Y pairing region (13). Clusters of minisatellites at the ends of the chromosomes do not apply for the mouse genome (14).

Techniques of Typing Minisatellites. As mentioned above, variation in the minisatellite pattern of length mutations can be studied by means of restriction

analyses and the use of probes, which can hybridize with a large number of minisatellite loci. The analytical technique has been further developed by the use of PCR amplification, giving an additional sensitivity. However, the disadvantage of PCR analysis of length variation is that many minisatellite alleles are too long for efficient amplification. Jeffreys et al. (9) have introduced a new PCR system, MVR, which has implied a solution of that problem and provided increased sensitivity by enabling analyses of internal and often subtle variations of internal repeats. The method is based on the use of primers, which are specific for repeat variants and which enable a successive PCR analysis of long stretches of repeated sequences with occasional variants. Such internal variation is present in almost all minisatellites.

Mutational Changes. The mutation frequency of minisatellites does not seem to be dependent on the length of the allele in the same way as in microsatellites (15). Short arrays of repeats can be stable over millions of years (16), while long alleles can have an extremely high frequency of mutational changes (up to 15%). The high instability of some human minisatellites seems to be a property of the repeated sequence itself, as indicated by the fact that the unstable human minisatellite MS1 retained its instability also after being inserted into the genome of yeast, *Saccharomyces* (17). In five highly unstable loci the rate of length-change mutations was related to their observed heterozygosity, indicating that the changes were selectively neutral.

Mutational changes of minisatellites are not randomly distributed, but occur predominantly at one end of the locus. This peculiar polarity was revealed by MVR-PCR analysis of three minisatellite loci (9). The occurrence of such polar hot spots was also found in pedigree analysis of germline mutations (18).

Different mechanisms of germline length changes of minisatellites can be visualized—replication slippage, intramolecular recombination, unequal sister chromatid exchange, and unequal interallelic recombination or gene conversion (10). The fact that no length change has been recorded involving an exchange of flanking markers eliminates a simple crossing-over model. About half of the length mutations recorded for three alleles studied by Jeffreys' group (18) were formed through small patch exchange between alleles, presumably involving gene conversionlike events (a process through which an allele in one of

the chromosomes is replaced by an allele in the other chromosome). Some mutations are of intraallelic origin. Anomalous repeats, not corresponding to either allele may have been brought about by mismatch repair.

Dubrova et al. (19) studied minisatellite length germline mutations in male mice induced by 0.5 or 1.0 Gy γ -radiation. The frequency of mutation was considerably higher than other end points, but the doubling dose effect was approximately the same. The data indicated that the selection against the mutations was insignificant.

Practical Application of Minisatellite Fingerprinting. The extreme individual variation of minisatellite and microsatellite pattern (below) has provided a new and exceedingly efficient tool for the recognition of individuals by their electrophoretic pattern. Already in the original fingerprint analysis the chance of two unrelated persons exhibiting the same pattern was extremely low—theoretically somewhere around 10^{-12} . Later methodological improvements have increased the sensitivity.

The possibility of amplifying DNA by PCR has made it possible to use extremely small material for minisatellite typing—single hairs or tiny blood stains. The fingerprinting of satellite DNA therefore has lent itself to analyses in forensic medicine and also in historical and archeological samples. The use of minisatellite fingerprinting in legal contexts has, however, caused much debate. Critical comments have emphasized the risk for deficient laboratory control, lack of clear definition of match of electrophoretic bands, dependence on the gel system used, and the question of statistical weight of apparent match between samples. Furthermore, it has been pointed out by the dominant critics Lewontin and Hartl (20) that error may be brought up by variation in allele frequency between subpopulations. A point of particular relevance in paternity establishment are germline mutations. Some prudence has been justified when introducing minisatellite fingerprinting, i.e., for forensic and legal purposes, but it seems that these possible sources of errors can be overcome and they are largely mitigated by the MVR-PCR typing system of both alleles for both length variation and internal variation (above). Although the reliability of the fingerprinting method has been questioned in some conspicuous legal cases, the use of this tool nevertheless has become more and more a routine procedure in forensic medicine.

The occurrence of minisatellites and other repetitive DNA sequences is not

restricted to humans and other mammals, but they have a wide distribution throughout the organism world. The use of minisatellite and microsatellite typing has become an important and highly valuable new tool in population ecology (21). Soon after the disclosure of the highly variable minisatellites by Jeffreys and co-workers in humans, investigations by Burke and Bruford (22) showed a pronounced variation in fingerprints within and between bird species. A population analysis of house sparrows through analyses of blood samples from each individual by fingerprinting demonstrated the usefulness of fingerprint mapping for analyses of population structure, mating selection, and various polygamous pairing strategies. Repetitive DNA sequences in sufficiently stable minisatellites and microsatellites are also of use in phylogenetic investigations of evolutionary processes by comparisons of species, subspecies, and populations.

Association of Minisatellites with Human Diseases. While no obvious evolutionary advantage of minisatellites can be discerned at the present state of our knowledge, there are some cases of pathogenic minisatellites. The best known case concerns the minisatellite connected with the Ha-ras protooncogene locus, *HRAS1* VNTR. This minisatellite is located 1000 bp downstream of the polyadenylation signal (23). It contains repeat units of 28 bp, forming about 30 alleles. Four of these, comprising 94% of the alleles, have given rise to the other alleles (24). The rarer alleles are three times more common in cancer patients than in controls and these alleles are associated with multiple forms of cancer. The data indicate that they contribute to 1 of 11 cases of cancer (25). The odds ratio for the association between the rare *HRAS1* minisatellite alleles and cancer were, according to Krontiris et al. (26), as shown in Table 1.

Concerning the mechanism behind the association between the *HRAS1* minisatellites and cancer two possibilities have been discussed (25,26). The rare alleles may exhibit a linkage with a potential disease

locus and these alleles would then just be markers for the risk of cancer. Considering the fact that the high-risk alleles derive from all the four common alleles and presumably from many ancestral chromosomes, this hypothesis is not likely. An alternative hypothesis is based on the finding that the *HRAS1* minisatellite binds to the *rel*/NF- κ B family of transcriptional regulatory factors (27,28). It is suggested that pathogenic minisatellite mutations may disrupt nonpathogenic interactions with *rel* proteins.

A somewhat similar pathogenic situation is indicated for minisatellite mutations linked to the insulin gene (*INS*). The minisatellite *INS* VNTR is located 600 bp upstream of the transcriptional start site (29). The minisatellite is composed of 14 bp repeat units arranged in three allelic classes with modal lengths of 600 (Class I), 1200 (Class II) and 2200 (Class III). The presence of Class I minisatellite is associated with a doubling of the relative risk for type I diabetes mellitus (IDDM) (25). At least six genes, *IDDM* 1 to 6, contribute to the risk for diabetes, and *IDDM*2 has been mapped to the *INS* VNTR minisatellite. The *INS* minisatellite, furthermore, binds to a specific transcription factor, Pur-1. However, in this case the high-risk allele exhibits a weaker transcriptional effect than the low-risk alleles. Nevertheless, the sequence composition of the individual repeat units, in addition to the total length of the minisatellite, governs the transcriptional response (30).

It is likely that other pathogenic effects of minisatellites will be revealed in the future. It can be mentioned now that a minisatellite upstream of the human immunoglobulin heavy-chain gene *IGH* enhancer may have a suppression effect on immunoglobulin gene expression by transcriptional control in the same way as *HRAS1* and *INS* minisatellites (31).

The minisatellites of the *HRAS1*, *INS* and *IGH* genes do not have any homologous counterpart in nonprimate genes and it is therefore unlikely that they constitute true transcriptional elements; rather, they are recent acquisitions (25). It is more likely that the variation of minisatellites sometimes produces products that interact with transcriptional factors and that, as long as the effect on transcription keeps within a narrow range, it will not be strongly selected against (25).

Conclusions. The wide occurrence of highly variable minisatellite sequences has provided indispensable tools in genetic linkage analyses, forensic medicine, pater-

nity determination and population ecology. Also, it can be foreseen that the use of minisatellites and other repeated DNA sequences will play an even more essential role in the future, both in research and for various practical applications. The reliability of the fingerprinting of minisatellites for legal purposes has been the subject of discussion and some controversy. However, the application of new PCR techniques, improved control of the laboratory procedures, and more experience with minisatellite patterns in subpopulations can be expected to remove many of the problems under discussion. The occurrence of pathogenic minisatellites has given another dimension to this field of research. The *HRAS1* minisatellite seems to be of major importance in the cancer panorama—at least 50,000 cases of cancer a year can be expected to depend on the rare alleles of this minisatellite (26). Another important finding is the connection between a minisatellite linked with the insulin locus *INS* and type I diabetes. A minisatellite linked to the enhancer of the immunoglobulin gene *IGH* is a third potentially important case. In all these cases, the effect of the minisatellites seems to occur through binding to transcription factors.

Microsatellites

Occurrence. Microsatellites are repetitive sequences of mostly 2 to 4 nucleotides with a widespread occurrence particularly in multicellular organisms. In the human genome, dinucleotide repeats occur on average every 30,000 bp and somewhat less frequently for the more complex units (32). These repeats therefore constitute a significant part of human DNA. Concerning the evolutionary significance of microsatellites, hardly anything but disadvantages can be discerned. Several human disorders have been attached to amplification of microsatellite sequences and other evidence of negative effects of microsatellites can be traced. Formation of tandem duplications of the short sequences that build up microsatellites can easily occur as an error during DNA replication, and further amplification through strand slippage can occur in successive DNA replication, giving rise to longer stretches of minisatellite repeats. An accumulation of dispersed repeated sequences of simple nucleotide units can be expected to imply an increased risk of homologous recombination between chromosomal segments and resulting in translocations, deletions, and inversions. Filamentous ascomycetes

Table 1. Odds ratio for the association between rare *HRAS1* minisatellite alleles and cancer.^a

Type of cancer	Odds ratio	p Value
Bladder	2.30	>0.001
Breast	1.68	0.001
Colorectal	2.21	0.002
Leukemia	2.29	0.001
Lung	1.55	0.051
Melanoma	1.56	0.091

^aData from Krontiris et al. (26).

such as *Neurospora crassa* do not seem to tolerate the burden that repetitive and apparently useless DNA implies. Presumably as a consequence, *Neurospora* has only 10% repetitive DNA as compared to 50% in higher organisms (33). To counteract the accumulation of dispersed homologous microsatellite sequences, these sequences are subjected to a high mutation rate through "repeat-induced point mutations" (RIP). All repeated sequences above 1 kilobase in *Neurospora* show signs of "RIPping." The primary function of RIP is to protect the organism not only against "parasite DNA" but also against viruses and transposons (33). Although it seems that higher organisms are less sensitive to these repeated sequences, there are reasons to believe that a similar protective device is operating also (below). At our present state of knowledge it is difficult to visualize any positive biological function at least for most microsatellites and it thus seems that they constitute true "parasite" or "selfish" DNA in the sense outlined by Orgel and Crick (1).

Although there are principal differences between micro- and minisatellites, the borderline between them is arbitrarily set on the bases of the length of the repeat units. From an evolutionary point of view, it is likely that minisatellites can be generated from microsatellites. Two hypotheses have been presented to account for the common core of minisatellites (34). According to a transposition model proposed by Jeffreys et al. (3), related core sequences between minisatellites are the result of transpositions mediated by sequences flanking the minisatellites VNTR. In support of this hypothesis, there are observations indicating an association of minisatellite VNTRs with dispersed repetitive elements such as human *Alu* and transposonlike sequences. Sequence divergence is brought about by subsequent mutational changes, which are carried to other repeats of the tandem array by unequal exchange. However, many minisatellites with related core sequences do not exhibit such an association with dispersed repeats flanking the tandem array (3,34), making it unlikely that they emanate from this kind of a transposition process. Another model, the expansion hypothesis, is based on the concept of core sequences containing motifs that enhance the expansion of tandem repeats independently at different loci (3). Short tracts of simple repeats would serve as the raw material for expansion by slipped strand mispairing into more complex minisatellites. This model would predict that one could

trace the development from microsatellites to minisatellites by "fossils" of microsatellites in close association or interdispersed with minisatellite VNTRs (34). Several examples of such an association have been recorded, indicating the generation of minisatellites from microsatellites.

Analytical Methods for Microsatellites.

Simple tandem repeat loci have been isolated from genomic libraries by hybridization screening, using relatively short oligonucleotide repeat sequences. However, the experience from isolation of minisatellite loci suggests that the use of long (>200 bp), tandemly repeated probes is more efficient than short probes to isolate longer tandem arrays; and longer probes would also better tolerate interspersed variant repeats. Armour et al. (35) therefore developed a more efficient system for the isolation of short repeats. Their system is based on a prior enrichment for tandemly repeated DNA fragments by hybridization to long tandemly repeated targets. A library of restriction fragments with appropriate linkers for PCR amplification is constructed. From amplified fragments of 400 to 1000 bp, tandem repeat-containing fragments are selected by hybridization to long arrays of either mixed trimeric or mixed tetrameric repeat sequences. Both natural and synthetic sequences were used as targets in the hybridization selection. This enrichment procedure enables a rapid isolation of a large number of microsatellite clones. In this way, Armour et al. isolated 46 tandem repeat arrays (27 tetrameric, 19 trimeric), which were sequenced and characterized (35).

Instability and Mutational Changes.

Many microsatellites are unstable—in some cases exceedingly so. In particular, CG-rich trinucleotide and CA dinucleotide repeats exhibit high instability and they are orders of magnitude more variable than other tandem repeats. The reason for this specificity in instability is not known. In extreme cases all cells in the organism have different lengths of the microsatellite (32). The instability is highly influenced by the length of the microsatellite with an increased instability with increasing length. CG-rich trinucleotides and CA dinucleotides form four groups (32):

- Short repeat length—stable
- Repeats of middle length—polymorphic, but stable between generations
- Long alleles—instability increased by several orders of magnitude, constituting "premutational alleles," which are not stable between generations

- Long and extremely unstable alleles, also exhibiting mitotic instability. For fragile X the likelihood of instability was as follows: below 50 repeats, no risk; around 60 repeats, low risk; 70 to 86 repeats, high risk; above 86 repeats, absolute likelihood (36).

The most common repeat length mutations involve only relatively small changes. *In vitro* studies have indicated that strand slippage during DNA replication constitutes the major cause of these length mutations (37). Furthermore, there is a connection between replication slippage and DNA repair, as is indicated by mutations in DNA repair, giving rise to increased instability. In *Saccharomyces*, mutations in mismatch repair genes increased replication slippage 100 to 700 times in poly-(GT) repeats (38). Infrequently, huge length increases occur, resulting in the extreme instability of the long alleles group above. This sudden increase in trinucleotide repeat length, which causes several human diseases, must involve some mechanism other than replication slippage. In cell culture 1000-fold amplification has been observed for the dihydrofolate reductase gene (39). This drastic amplification involves an episomal mechanism. The gene is excised and copied, presumably by a rolling circle process, and reintegrated into nonhomologous chromosomal sites. This mechanism, however, is not the likely one to explain the amplification of trinucleotide repeats. Unlike the case with the episomal mechanism, the amplification of trinucleotides never involves surrounding DNA and it always occurs *in situ*. A model to explain the expansion of trinucleotides (32) takes into consideration the difficulty in replicating CG-rich sequences by polymerases (36). It is possible that replication of these repeats gives rise to premature termination and reinitiation events, generating multiple incomplete strands. Extensive increases in length can then be induced by a strand switching between the incomplete strands. This model predicts that an increased rate and an increased length of the expansion will occur with increasing initial length of the trinucleotide sequence. These predictions have been experimentally observed (36).

Repeat-induced Point Mutations. The accumulation of repeated sequences, particularly of microsatellites, implies a risk for homologous recombination between dispersed repeats, causing translocation and other chromosomal aberrations. Fungi like *Neurospora* (above) are less tolerant towards such repeated sequences than

higher multicellular organisms, and they have developed defense mechanisms against homologous repeated deletions by RIP (33). RIP recognizes duplicated sequences and induces G.C to A.T mutation. This mutation is associated with methylation of cytosin and a high frequency of recombination between tandem repeats. Both alleles are mutated in this process, and the genetic mechanism seems to be comparable to the recognition of homologous sequences by recombination processes. At the molecular level, the high frequency of G.C to A.T transitions is probably caused by enzymatic deamination of cytosin or 5-methylcytosine. The normal repair of these lesions may be turned off in ascogenous tissue or overwhelmed by RIP. This mutational process may be an integral part of "genome cleaning" during the period between fertilization and karyogamy in fungi, which also includes a high frequency of intrachromosomal recombination, deleting tandemly repeated genes. The sequence divergence by RIP can be sufficient to prevent recognition of homology and subsequent recombination between dispersed DNA regions. This form of genetic instability potentially stabilizes the gross organization of the genome (33).

Another form of RIPping has been described by Rand (40) in the mitochondrial DNA in crickets (*Gryllus*). A repeated sequence of 220 bp in length was found to be a hot spot for point mutations, deletions, and insertions. The mutational changes were localized in and around a 14 bp G.C-rich sequence. This mutation process apparently involves a mechanism other than the RIPing in *Neurospora*, as neither methylation nor a bias towards G.C to A.T mutations was observed in the cricket mtDNA. As it is not clear if these mutations are induced by the repeats or associated with the repeats, this process has been named repeat associated point mutation (RAP).

The potential genetic and biological disadvantages of repeated sequences, such as microsatellites, can be expected to be of general relevance, although the threshold for negative effects presumably is higher in higher organisms. Kricker et al. (41) have

pointed out that vertebrate chromosomes would be threatened by illegitimate recombination between repeated sequences, such as mobile elements and pseudogenes. To counteract this "genetic time bomb," a strategy based on methylation and associated mutations through methylation and deamination of 5-methylcytosine in *CpG* has been developed.

Instability and Mismatch Repair. The stability of microsatellites is dependent on an intact mismatch DNA repair. The loss of this repair function in *Saccharomyces* increased the instability of microsatellites drastically (above). The data on yeast indicated that the strong effect on the stability of poly (GT) recorded depended on errors in the excision of mismatch bases after DNA slippage, most of which are corrected by mismatch repair in wild-type cells (38). The discovery of a similar case with colon cancer has attracted much attention. Fifteen percent of colorectal cancers have a hereditary background, hereditary non-polyposis colon cancer (HNPCC). One gene involved in this cancer was localized to chromosome 2 and linked to this locus was a microsatellite with an array of AC repeats. In the tumors of HNPCC patients, mutations in this gene caused an extensive instability, not only of the AC repeats linked to the gene, but also of microsatellites elsewhere in the genome, which were subjected to thousands of changes (42,43). The gene in chromosome 2, responsible for the genetic instability of HNPCC tumors, was homologous to the mismatch repair gene *MutS* in *E. coli* and *MSH2* in yeast (44,45). Subsequently, three more human genes, homologous with the mismatch repair genes in *E. coli* and yeast, have been linked to HNPCC (46) (Table 2). Parsons et al. (47) recently reported a subset of HNPCC patients with a high frequency of microsatellite mutations not only in their tumors but also in nonneoplastic cells. These patients furthermore had very few tumors, showing that deficient mismatch repair and succeeding mutations can be

compatible with normal development and not sufficient for tumor development. On the other hand, instability of microsatellites is also generated by mechanisms other than deficient mismatch repair genes—for example, deficiency in exonuclease (48). Several cancer forms have been found to be associated with microsatellite instability; these include gastric, pancreatic endometrial, Barrett's esophageal, and lung cancer (49).

Association of Microsatellites with Human Diseases. The previous section dealt with microsatellite instability in connection with DNA repair deficiency and the association of this instability with cancer. This association between microsatellite instability and the disease is not a causal one, but presumably a common result of the lack of mismatch repair of DNA. However microsatellites have attracted a great deal of attention in recent years because of a direct connection between expanded arrays of CG-rich trinucleotides and several human neurological diseases. Table 3 shows five diseases of trinucleotide reiteration that have been characterized (50).

The microsatellite sequence in these diseases are linked to a coding gene, which is affected by the expansion of the trinucleotide sequence. These cases of microsatellite-dependent disease represent two classes. Fragile X and myotonic dystrophy have their trinucleotide sequence linked to the noncoding ends of the gene, while in the three diseases with CAG expansion, coding for polyglutamine, the microsatellite is located within the coding part of the gene. An initial increase of the trinucleotide sequence functions as a pre-mutational event. Above a critical number of repeats, the system becomes unstable and usually more sequences are added, eventually resulting in symptoms. Another characteristic of these diseases is the fact that the symptoms tend to be more severe in subsequent generations because of amplification during gametogenesis or in the zygote, a process named genetic anticipation. This anticipation is sex linked and *inter alia*

Table 2. Mismatch repair genes.^a

<i>E. coli</i>	<i>S. cerevisiae</i>	Homo	% HNPCC
<i>MutS</i>	<i>MSH2</i>	<i>hMSH2</i>	<50
<i>MutL</i>	<i>MLH1</i>	<i>hMLH1</i>	<30
<i>MutL</i>	<i>PMS1</i>	<i>hPMS1</i>	Few cases
<i>MutL</i>		<i>hPMS2</i>	Few cases

^aData from Fishel and Kolodner (46).

Table 3. Diseases associated with trinucleotide reiteration.^a

Disease range	Reiterated trinucleotide	Normal range of reiteration	Disease range of reiteration
Spinal and bulbar muscular atrophy, Kennedy's disease	CAG	11–33	40–62
Huntington's disease	CAG	11–34	42–100
Spinocerebellar ataxia type 1	CAG	<29–36	43–60
Fragile X	CGG	6–54	250–4000
Myotonic dystrophy	CTG	5–30	>50

^aData from Green (50).

occurs through the mother in fragile X and myotonic dystrophy but through the father in Huntington's disease. This process of anticipation is connected with methylation of cytosine and genetic imprinting.

Fragile X was the first recognized case of a genetic disease with an instability of trinucleotide repeats. It is a common neurological disease that causes mental retardation, which is inherited as an X-linked dominant trait. It is manifested by chromosome breakages at specific sites. The disease is associated with an expansion of a microsatellite repeat of CGG trinucleotides in the 5' untranslated end of the gene *FMRI*. This leads to a hypermethylation of the promoter region and a down regulation of the gene expression. The mechanism of inactivation of neighboring loci by non-coding repeats has a counterpart in the heterochromatization of euchromatin by tandem repeats as observed in *Arabidopsis* and in *Drosophila* (51). Although the instability of the trinucleotide repeats depends on the length of the sequence, the length at which an instability of the microsatellite begins to occur varies between 35 and 55 repeats. This long "gray zone" was shown by Eichler et al. (52) to depend on the interspersions by AGG trinucleotides. Most alleles of CGG repeats contain two AGGs and the instability depends on the number of uninterrupted CGG repeats. The uninterrupted length under which the microsatellite is stable turned out to be 34 to 37 CGGs, which is in agreement with the corresponding number in other triplet repeat diseases including myotonic dystrophy, Kennedy's disease, Huntington's disease, spinocerebellar ataxia, and dentatorubral pallidoluyisian atrophy. The loss of AGG causes an increased uninterrupted length of CGG sequences and is therefore probably an important mutational event for the predisposition to fragile X. The mechanism of expansion of the trinucleotide sequence presumably rests on slippage during DNA replication. To explain the rapid expansion of a large sequence of triplets, Eichler et al. (52) argue in favor of a slippage mechanism dependent on the lagging and leading strand. Based on the observed polarity of expansions at the 3' end, they propose a slippage process involving a whole Okasaki fragment, spanning 150 to 200 bp within trinucleotide repeat alleles of about 70 CGGs (210 bp). Concerning the non-mendelian increase of the effect of mutated genes from one generation to the next, this "anticipation" has been reported as a postzygotic process in fragile X syndrome,

while premutational increase takes place during meiosis (51).

Myotonic dystrophy, another neurological disease, depends on an expansion of a sequence of CTG at the 3' untranslated end of the gene myotonic dystrophy protein kinase (MDPK). The expanded trinucleotide sequence eliminates transcription of MDPK. Above a threshold of about 146 bp no mRNA for MDPK could be observed (53). An increased nucleosome binding of such expanding repeats, leading to a transcriptional repression, has been proposed as a mechanism (53). It has further been shown that the increased nucleosome binding exerts an effect on the post-transcriptional processing of the transcript from expanded alleles but not on the initiation of the transcription (54). The threshold for the symptoms of myotonic dystrophy of 146 bp corresponds with the DNA length for a nucleosome (53).

Huntington's disease is an autosomal neurodegenerative disorder that depends on an expansion of CAG tandem repeats, giving rise to polyglutamin. This microsatellite is located within the coding region of the Huntington's disease gene, HD or IT15, which codes for the protein huntingtin. Although the disease is dependent on the stretch of CAG repeats, the instability giving rise to an expansion of the polyglutamine array seems to be influenced by another trinucleotide repeat of CCG downstream of the CAG repeat, coding for proline (55). Huntington's disease usually has a late onset, but with increasing expansion of the trinucleotide repeats, due to "anticipation" through male gametes, the symptoms become more severe and onset earlier in subsequent generations. The function of huntingtin is not known and its expression is similar in patients and controls. The length expansion makes the protein not merely useless, but actively harmful by a gain of function. Zeitlin et al. (56) showed that huntingtin is indispensable, as null mutation of the huntingtin gene in mice caused death of the embryo. The data on mice further suggested that huntingtin is involved in counterbalancing programmed cell death, apoptosis (56). Reports by Li et al. (57) indicate that the pathological effects by the expansion of the CAG repeats depend on interaction between huntingtin and other cellular proteins. They identified a protein, huntingtin-associated protein, HAP-1, that binds to huntingtin. This binding is enhanced by an expanded polyglutamine. The HAP-1 protein is enriched in the

brain, which may explain the localized effect of the disease to brain tissue.

Spinal and bulbar muscular atrophy, Kennedy's disease, is an X-linked disease and the only polyglutamine-dependent neurological disorder for which the function of the protein involved is known. It constitutes the androgen receptor (AR), which is a ligand-activated transcription factor. The AR contains, in the coding region, the polyglutamine tract by CAG repeats. As in the other microsatellite-dependent diseases, the severity of the disease is correlated with the expansion of the microsatellite. Chamberlain et al. (58) showed that progressive expansion of the polyglutamine tract in human AR caused a linear decrease in the binding of the receptor to androgen and a decrease in activating transcription of AR-responsive genes. However, the data indicated that there was a threshold, as the expansion of the trinucleotides did not completely eliminate AR activity, and that the residual activity was sufficient to develop male primary and secondary sex characteristics.

Inactivation Mechanisms by Polyglutamine. Many data on the relationship between expansion of trinucleotides and neurological disorders are now available, but the mechanistic cause of the effects of the expanded repeats at a molecular level is not clear. An attractive possibility is that long trinucleotide repeats confer structural changes of DNA, and that these changes constitute the ultimate reason for the pathological behavior. Yano-Yanagisawa et al. (59) found in the mouse brain two trinucleotide repeat-binding proteins—TRIP-1 and TRIP-2—which bind specifically to repeats of AGC, AGT, GGC, and GGT, but no other trinucleotides. The AGC-repeat binding activity is of interest concerning polyglutamine. (CAG)-repeats were found to contain clusters of non-B DNA structural units, formed by each AGC trinucleotide repeating unit: 5'.....(C AG)(C AG)(C AG)(C AG).....3'. In non-B DNA, cytosines are specifically base unpaired. The property of trinucleotides to adopt an unusual DNA structure may contribute to their abnormal behavior. Recently Gacy et al. (60) presented data suggesting that hairpin formation in microsatellite repeated sequences may provide a common explanation for several characteristics of simple nucleotide repeat expansion and pathological effects. Hairpin formation and stability are correlated with the length of the repeat sequence. They would, for example, explain the stabilizing effect of AGG punctuation

on *FMRI* in Fragile X (above) on the basis of its interruption of hairpin stability. The repeats that form hairpin structures would disrupt normal DNA replication. Above a critical threshold length, stable hairpin structures are formed, leading to replication errors and further expansion.

There are other neurological diseases with characteristics that resemble the ones established as dependent on polyglutamine expansion, such as the phenomenon of anticipation. It is therefore likely that more such neurological disorders will be added to this class of polyglutamine degenerative diseases. The discovery of proteins that interact with polyglutamine stretches with an intensity dependent on the length of CAG repeats opens new possibilities for identifying other disorders of this type. Trotter et al. (61) have characterized a monoclonal antibody that selectively recognizes polyglutamine expansions in Huntington's disease, spinocerebellar ataxia SCA1 and Machado-Joseph disease SCA3—all known glutamine-repeat disorders. An expansion of polyglutamine was detected in this way with spinocerebellar ataxia SCA2 and the dominant cerebellar ataxia with retinal degeneration. There are indications that schizophrenia and bipolar disorders may belong to the same group of diseases. O'Donovan et al. (62) found that schizophrenia patients and patients suffering from bipolar disorders had expanded trinucleotide sequences of CAG and its complement CTG as compared to controls. The connection between expanded trinucleotide CAG repeats and degenerative disorders may open the possibility in the future to design therapeutic molecules that would interfere with abnormal polyglutamine stretches.

Telomeres

The ends of the chromosomes in most organisms consist of a tandem array of

simple DNA sequences, constituting the telomeres [Zakian (63) presents a recent review]. This arrangement of the chromosome ends is dictated by the fact that the DNA polymerase cannot reproduce both DNA strands to the ends without losing the tip sequence of one of the strands. Therefore the tip of the chromosomes is organized with noncoding repeated sequences, which can be lost without losing coding DNA [Ligner et al. (64) present a recent overview]. The telomeres can be replaced by a protein-RNA enzyme, telomerase. Most telomeric repeat sequences are short, usually 5 to 8 bp, in mammals TTAGGG. *Drosophila* has an exceptional telomere structure without the short conservative repeats of the telomeres of most other organisms (65). Instead, *Drosophila* has one or more elements like long interspersed elements (LINE) mobile elements, and the replacement of the telomeres occurs by transposition of the telomere sequence to the chromosome ends. Proximal to the LINE sequences in *Drosophila* there is a sequence of tandem repeats, which probably are analogous to the subterminal middle repetitive regions, telomere-associated (TA) DNA, in other eukaryotes. The array of TA can expand or contract by means of a recombination mechanism. The composition of telomere-repeated sequences varies considerably among organisms, and evidently the telomere function does not require a specific DNA sequence. The DNA strand, which runs from 5' to 3' towards the end, has regularly more G residues, arranged in clusters, than the other strand. At least in ciliated protozoans, such as *Tetrahymena* and *Oxytricha*, and in yeast the G strand is extended to form a single-strand G tail. The G strand can form non-Watson-Crick base pairing structures, such as four-stranded helices and multiple G-G base

pairs. It is possible that this property is essential for the bouquet stage, formed by the telomeres during meiosis.

The function of the telomeres is to protect the ends of the chromosomes, not only from losing genetic material at each cell division, but also to prevent the ends from fusing with each other. As was shown by McClintock (66), broken chromosome ends fuse with each other, forming dicentric bridges and a breakage-fusion-bridge cycle. Because of the loss of DNA at the chromosome ends the telomeres have to be replicated in another way than the rest of the chromosomes. This replication is acquired by telomerase, which has a unique composition of protein and an RNA component. The protein part consists of two subunits in *Tetrahymena* (67). The replication of the telomere occurs by means of reverse transcriptase from the RNA component. In *Drosophila* this replication is performed through transposition of the telomere sequence (above). In humans most somatic tissues lose their telomerase activity and consequently the chromosome ends will shorten at each cell division, leading to the eventual death of the cell. This has led to the hypothesis that the telomere length functions as a biological clock, resulting in a programmed cell death after a certain number of cell divisions (68). In actual measurement of the telomerase activity in normal and immortal cancer cells telomerase activity was invariably repressed in normal somatic cells but was reactivated in various cancer cells (69). The important observation that the immortality of malignant cells is associated with telomerase activity has led to speculation that the telomere might constitute a target for cancer therapy. Similar speculations can also be applied concerning prevention of aging by reactivation of telomerase activity.

REFERENCES

1. Orgal LE, Crick FHC. Selfish DNA: the ultimate parasite. *Nature* 284:604-607 (1980).
2. Schimke RT. Gene amplification, drug resistance and cancer. *Cancer Res* 44:1735-1742 (1984).
3. Jeffreys AJ, Wilson V, Thein SL. Hypervariable minisatellite region in human DNA. *Nature* 314:67-73 (1985).
4. Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E et al. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 235:1616-1622 (1987).
5. Braman J, Barker D, Schumm J, Knowlton R, Donis-Keller H. Characterization of very highly polymorphic RFLP probes. *Cytogenet Cell Genet* 40:589 (1985).
6. Jeffreys AJ. 23rd Colworth Medal Lecture. Highly variable minisatellites and DNA fingerprint. *Biochem Soc Trans* 15:309-317 (1987).
7. Jeffreys AJ, Royle NJ, Wilson V, Wong Z. Spontaneous mutation rates to new length alleles at tandem repetitive hypervariable loci in human DNA. *Nature* 332:278-281 (1988).
8. Jeffreys AJ, Wilson V, Thein SL. Individual-specific 'fingerprints' of human DNA. *Nature* 316:76-79 (1985).
9. Jeffreys AJ, MacLeod A, Tamaki K, Neil DL, Monckton DG. Minisatellite repeat coding as a digital approach to DNA typing. *Nature* 354:204-209 (1991).
10. Armour JAL, Jeffreys AJ. Biology and application of human minisatellite loci. *Curr Opin Genet Dev* 2:850-856 (1992).

11. Hultén M. Chiasmata distribution in the normal human male. *Hereditas* 76:55–78 (1974).
12. Laurie DA, Hultén M. Further studies on chiasmata distribution and interference in the human male. *Ann Hum Genet* 49:203–214 (1985).
13. Cooke HJ, Brown WR, Rappold GA. Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. *Nature* 317: 687–692 (1985).
14. Julier C, DeGouyon B, Georges M, Guenet J-L, Nakamura Y, Avner P, Lathrop M. Minisatellite linkage maps in the mouse by cross-hybridization with human probes containing tandem repeats. *Proc Natl Acad Sci USA* 87:4585–4589 (1990).
15. Jeffreys AJ, Tamaki K, MacLeod A, Monckton DG, Neil DL, Armour JAL. Complex gene conversion events in germline mutation at human minisatellites. *Nat Genet* 6:136–145 (1994).
16. Gray IC, Jeffreys AJ. Evolutionary transience of hypervariable minisatellites in man and the primates. *Proc R Soc London B Biol Sci* 243:241–253 (1991).
17. Cederberg H, Agurell E, Hedenskog M, Rannug U. Amplification and loss of repeat units of the human minisatellite MS1 integrated into chromosome III of haploid yeast strain. *Mol Gen Genet* 238:38–42 (1993).
18. Armour JAL, Monckton DG, Neil DL, Tamaki K, MacLeod A, Allen M, Crosier M, Jeffreys AJ. Mechanisms of mutation at human minisatellite loci. In: *Genome Rearrangement and Stability* (Davies KE, Warren ST, eds). *Genome Analysis* 7:43–57 (1993).
19. Dubrova YE, Jeffreys AJ, Malashenko AM. Mouse minisatellite mutations induced by ionizing radiation. *Nat Genet* 5:92–94 (1993).
20. Lewontin RC, Hartl DL. Population genetics in forensic DNA typing. *Science* 254:1745–1750 (1991).
21. Bruford MW, Wayne RK. Microsatellites and their application to population genetic studies. *Curr Opin Genet Dev* 3:939–943 (1993).
22. Burke T, Bruford MW. DNA fingerprinting in birds. *Nature* 327:149–152 (1987).
23. Capon DJ, Chen EY, Levinson AD, Seeburg PH, Goedder DV. Complete nucleotide sequence of the *T24* human bladder carcinoma oncogene and its normal homologue. *Nature* 302:33–37 (1983).
24. Kasperczyk A, DiMartino NA, Krontiris TG. Minisatellite allele diversification: the origin of rare alleles of the *HRAS1* locus. *Am J Hum Genet* 47:854–859 (1990).
25. Krontiris TG. Minisatellites and human disease. *Science* 269:1682–1683 (1995).
26. Krontiris TG, Derlin B, Karp DD, Robert NJ, Risch N. An association between the risk of cancer and mutations in the *HRAS1* minisatellite locus. *N Engl J Med* 329:517–523 (1993).
27. Trepicchio WL, Krontiris TG. Members of the rel/NF- κ B family of transcriptional regulatory factors bind the *HRAS1* minisatellite DNA sequence. *Nucleic Acids Res* 20:247–2434 (1992).
28. Green M, Krontiris TG. Allelic variation of reporter gene activation by the *HRAS1* minisatellite. *Genomics* 17:429–434 (1993).
29. Bell GI, Selby M, Rutter WJ. The highly polymorphic region near the human insulin gene is composed of simple tandemly repeated sequences. *Nature* 295:31–35 (1982).
30. Kennedy GC, German MS, Rutter WJ. The minisatellite in the diabetes susceptibility locus *IDDM2* regulates insulin transcription. *Nat Genet* 9:293–258 (1995).
31. Trepicchio WL, Krontiris TG. IGH minisatellite suppression of USF-binding-site and E μ -mediated transcriptional activation of the adenovirus major late promoter. *Nucleic Acids Res* 21:977–985 (1993).
32. Kuhl DPA, Caskey CT. Trinucleotide repeats and genome variation. *Curr Opin Genet Dev* 3:404–407 (1993).
33. Selker EU. Premeiotic instability of repeated sequences in *Neurospora crassa*. *Annu Rev Genet* 24:579–613 (1990).
34. Wright JM. Are minisatellites the evolutionary progeny of microsatellites? *Genome* 37:34335–34346 (1994).
35. Armour JAL, Neumann R, Gobert S, Jeffreys AJ. Isolation of human simple repeat loci by hybridization selection. *Hum Mol Genet* 3:599–605 (1994).
36. Fu Y-H, Kuhl DP, Pizzuti A, Pieretti M, Sutcliffe JS, Richards S, Verkerk AJ, Holden JJ, Fenwick RG Jr, Warren ST et al. Variation of the CGG repeat at the fragile X site resulting in genetic instability: resolution of the Sherman paradox. *Cell* 67:1047–1058 (1991).
37. Schlötterer C, Tautz D. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res* 20:211–215 (1992).
38. Strand M, Prolla TA, Liskay RM, Petes TD. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* 365:274–276 (1993).
39. Trask BJ, Hamlin JL. Early dihydrofolate reductase gene amplification events in CHO cells usually occur on the same chromosome arm as the original locus. *Genes Dev* 3:1913 (1989).
40. Rand DM. RIPping and RAPPING at Berkeley. *Genetics* 132:1223–1224 (1992).
41. Kricker MC, Drake JW, Radman M. Duplication-targeted DNA methylation and mutagenesis in the evolution of eukaryotic chromosomes. *Proc Natl Acad Sci USA* 89:1075–1079 (1992).
42. Peltomäki P, Aaltonen LA, Sistonen P, Pylkkanen L, Mecklin JP, Jarvinen H, Green JS, Jass JR, Weber JL, Leach FS et al. Genetic mapping of a locus predisposing to human colorectal cancer. *Science* 260:810–812 (1993).
43. Aaltonen LA, Peltomäki P, Leach FS, Leach FS, Sistonen P, Pylkkanen L, Mecklin JP, Jarvinen H, Powell SM, Jen J et al. Clues to the pathogenesis of familial colorectal cancer. *Science* 260:812–816 (1993).
44. Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J, Kane M, Kolodner R. The human mutator gene homolog *MSH2* and its association with hereditary nonpolyposis colon cancer. *Cell* 75:1027–1038 (1993).
45. Leach FS, Nicolaides NC, Papadopoulos N, Liu B, Jen J, Parsons R, Peltomäki P, Sistonen P, Aaltonen LA, Lystrom-Lahti M et al. Mutations of a *MutS* homolog in hereditary non-polyposis colorectal cancer. *Cell* 75:1215–1225 (1993).
46. Fishel R, Kolodner RD. Identification of mismatch repair genes and their role in the development of cancer. *Curr Opin Genet Dev* 5:382–395 (1995).
47. Parsons R, Li G-M, Longley M, Modrich P, Liu B, Berk T, Hamilton SR, Kinzler KW, Vogelstein B. Mismatch repair deficiency in phenotypically normal human cells. *Science* 268:738–740 (1995).
48. Johnson RE, Kovvali GK, Prakash L, Prakash S. Requirement of the yeast RTH1 5' to 3' exonuclease for the stability of simple repetitive DNA. *Science* 269:238–240 (1995).
49. Meltzer SJ, Yin J, Manin B, Rhyu MG, Cottrell J, Hudson E, Redd JL, Krasna MJ, Abraham JM, Reid BJ. Microsatellite instability occurs frequently and in both diploid and aneuploid cell populations of Barrett's-associated esophageal adenocarcinoma. *Cancer Res* 54:3379–3382 (1994).
50. Green H. Human genetic diseases due to codon reiteration: relationship to an evolutionary mechanism. *Cell* 74:955–956 (1993).
51. Tautz D, Schlötterer C. Simple sequences. *Curr Opin Genet Dev* 4:832–837 (1994).
52. Eichler EE, Holden JJ, Popovich BW, Reiss AL, Snow K, Thibodeau SN, Richards CS, Ward PA, Nelson DL. Length of uninterrupted CGG repeats determines instability in the *FMR1* gene. *Nat Genet* 8:88–94 (1994).
53. Wang Y-H, Amirhaeri S, Kang S, Wells RD, Griffith JD. Preferential nucleosome assembly at DNA triplet repeats from myotonic dystrophy gene. *Science* 265:669–671 (1994).
54. Krahe R, Ashizawa T, Abbruzzese C, Roeder E, Carango P, Giacanelli M, Funanage VL, Siciliano MJ. Effect of myotonic dystrophy trinucleotide repeat expansion on DMPK transcription and processing. *Genomics* 28:1–14 (1995).
55. Andrew SE, Goldberg YP, Theilmann J, Zeisler JR, Hayden

- MR. ACCG repeat polymorphism adjacent to the CAG repeat in the Huntington disease gene: implication for diagnostic accuracy and predictive testing. *Hum Mol Genet* 3:65–67 (1994).
56. Zeitlin S, Liu J-P, Chapman DL, Papaioannou VE, Efstradiadis A. Increased apoptosis and early embryonic lethality in mice nullizygous for the Huntington's disease gene homologue. *Nat Genet* 11:155–163 (1995).
57. Li X-J, Li S-H, Sharp AH, Nucifora FC Jr, Schilling G, Lanahan A, Worley P, Snyder SH, Ross CA. A Huntington-associated protein enriched in brain with implication for pathology. *Nature* 378:398–402 (1995).
58. Chamberlain NL, Driver ED, Mjesfeld RL. The length and location of CAG trinucleotide repeat in the androgen receptor N-terminal domain affect transactivation function. *Nucleic Acids Res* 22:3181–3186 (1994).
59. Yano-Yanagisawa H, Li Y, Wang H, Kohwi Y. Single-stranded DNA binding proteins isolated from mouse brain recognize specific trinucleotide repeat sequences *in vitro*. *Nucleic Acids Res* 23:2654–2660 (1995).
60. Gacy AM, Goellner G, Juranic N, Macura S, McMurray CT. Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*. *Cell* 81:533–540 (1995).
61. Trotter Y, Lutz Y, Stevanin G, Imbert G, Devys D, Cancel G, Saudou F, Weber C, David G, Tora L et al. Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias. *Nature* 378:403–406 (1995).
62. O'Donovan MC, Guy C, Craddock N, Murphy KC, Cardno AG, Jones LA, Owen MJ, McGuffin P. Expanded CAG repeats in schizophrenia and bipolar disorders. *Nature Genet* 10:380–381 (1995).
63. Zakian VA. Telomeres: beginning to understand the end. *Science* 270:1601–1607 (1995).
64. Lingner J, Cooper JP, Cech TR. Telomerase and DNA end replication: no longer a lagging strand problem? *Science* 269:1533–1534 (1995).
65. Mason JM, Biessmann H. The unusual telomeres of *Drosophila*. *Trends Genet* 11:58–62 (1995).
66. McClintock B. The stability of broken ends of chromosomes in *Zea mays*. *Genetics* 26:234–282 (1941).
67. Collins K, Kobayashi R, Greider CW. Purification of *tetrahymena* telomerase and cloning of genes encoding the two protein components of the enzyme. *Cell* 81:677–686 (1995).
68. Harley CB, Futcher, Greider CW. Telomeres shorten during ageing of human fibroblasts. *Nature* 345:458–460 (1990).
69. Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PL, Coviello GM, Wright WE, Weinrich SL, Shay JW. Specific association of human telomerase activity with immortal cells and cancer. *Science* 266:2011–2015 (1994).